



Teaching Guide

| Identifying Data | | | | | 2023/24 |
|---------------------|---|--------|-------------------------|-----------|---------|
| Subject (*) | Omics Data Management and Modeling | | Code | 614G02042 | |
| Study programme | Grao en Ciencia e Enxeñaría de Datos | | | | |
| Descriptors | | | | | |
| Cycle | Period | Year | Type | Credits | |
| Graduate | 2nd four-month period | Fourth | Optional | 6 | |
| Language | Spanish | | | | |
| Teaching method | Face-to-face | | | | |
| Prerequisites | | | | | |
| Department | Ciencias da Computación e Tecnoloxías da Información | | | | |
| Coordinador | Fernández Lozano, Carlos | E-mail | carlos.fernandez@udc.es | | |
| Lecturers | Fernández Lozano, Carlos | E-mail | carlos.fernandez@udc.es | | |
| Web | cafernandezlo.github.io/es_github_cafernandezlo/teaching/ | | | | |
| General description | <p>Ao longo dos últimos anos, a cantidade de datos biolóxicos dispoñibles está a crecer de maneira exponencial. Isto fai posible que diferentes disciplinas científicas poidan dispoñer deles para estudar a nivel molecular diferentes organismos co obxectivo de xerar novo coñecemento e/o verificar o existente. O perfil dun investigador que utilice datos ómicos debe ser computacional pero, á súa vez, debe levar asociado un perfil de interese biolóxico para que as técnicas de análises a aplicar e os resultados obtidos cobren sentido. Devanditos coñecementos biolóxicos integranse durante as clases de forma sinxela para que o alumnado adquira as competencias necesarias no desenvolvemento das análises.</p> <p>A redución de custos e o incremento da capacidade de cómputo dos últimos tempos logrou achegar as plataformas de secuenciación masiva ao día a día da investigación e, nalgúns casos, á asistencia clínica diaria. Como exemplo, a día de hoxe, existen paneis ómicos que se utilizan para determinar se unha determinada paciente sofre un cancro de mama, o estadio no que se atopa, o subtipo concreto que padece e xéranse modelos de predición de recidivas e de resposta a fármaco. Todo iso é posible grazas aos datos ómicos e a análises bioinformáticos.</p> <p>Grazas ás aproximacións de ciencia de datos é posible comprender o funcionamento de fenómenos biolóxicos complexos a diferentes niveis (xenoma, proteoma, microbioma, etc) e é necesario aplicar novos paradigmas capaces de xestionar e analizar o enorme volume de datos do que se dispón.</p> | | | | |

Study programme competences

| Code | Study programme competences |
|------|--|
| A2 | CE2 - Capacidade para resolver problemas matemáticos, planificando a súa resolución en función das ferramentas dispoñibles e das restricións de tempo e recursos. |
| B2 | CB2 - Que os estudantes saiban aplicar os seus coñecementos ao seu traballo ou vocación dunha forma profesional e posúan as competencias que adoitan demostrarse por medio da elaboración e defensa de argumentos e a resolución de problemas dentro da súa área de estudo |
| B3 | CB3 - Que os estudantes teñan a capacidade de reunir e interpretar datos relevantes (normalmente dentro da súa área de estudo) para emitir xuízos que inclúan unha reflexión sobre temas relevantes de índole social, científica ou ética |
| B4 | CB4 - Que os estudantes poidan transmitir información, ideas, problemas e solucións a un público tanto especializado como non especializado |
| B7 | CG2 - Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables. |
| B8 | CG3 - Ser capaz de manter e estender formulacións teóricas fundadas para permitir a introdución e explotación de tecnoloxías novas e avanzadas no campo. |
| B9 | CG4 - Capacidade para abordar con éxito todas as etapas dun proxecto de datos: exploración previa dos datos, preprocesado, análise, visualización e comunicación de resultados. |
| B10 | CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións. |



| | |
|----|---|
| C1 | CT1 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida. |
| C4 | CT4 - Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade. |

| Learning outcomes | | | | |
|---|--|-----------------------------|---|----------|
| Learning outcomes | | Study programme competences | | |
| Coñecemento e aplicación de técnicas de análises de datos ómicos desde a xeración de datos científicos de secuencias de ácidos nucleicos ou secuencias de proteínas ata o estudo de diferenzas entre subtipos dunha determinada patoloxía | | A2 | B2 B3 B4 B7 B8 B9 B10 | C1 C4 |
| Manexar as ferramentas e contornas máis actuais no ámbito de análise de datos biolóxicos | | A2 | B9 B10 | C1 C4 |

| Contents | |
|---|--|
| Topic | Sub-topic |
| 1. Introducción aos datos ómicos | 1.1. O ADN 1.2. O dogma central da bioloxía molecular 1.3. As ómicas |
| 2. Traballo con secuencias moleculares | 2.1. O formato FASTA 2.2. Bases de datos abertas 2.3. Ferramentas para a análise de secuencias moleculares: BLAST, Clustal, Galaxy... |
| 3. Tecnoloxías de secuenciación masiva (NGS) | 3.1. A orixe coa secuenciación Sanger 3.2. Novas tecnoloxías NGS 3.3. Illumina, PacBio, MinION, Solexa 3.4. Diferenzas entre plataforma de secuenciación curta e longa, aplicacións máis frecuentes |
| 4. Análise da calidade e filtrado de secuencias | 4.1. Formato FASTAQ 4.2. Control e avaliación de calidade das secuencias 4.3. Filtrado das secuencias |
| 5. Ensamblaxe de xenomas e metaxenomas | 5.1. Ensamblaxe de xenomas 5.2. Ensamblaxe de novo 5.3. Ensamblaxe contra xenoma de referencia 5.4. Ferramentas software de ensamblaxe 5.5. Mapeado contra xenoma 5.6. Anotación de secuencias 5.7. Exemplos de uso |
| 6. Análise de expresión xénica mediante RNA-Seq | 6.1. Preprocesado 6.2. Análise de expresión diferencial con R/Bioconductor: DESeq2 6.3. The Cancer Genome Atlas (TCGA) |
| 7. Análise do metaxenoma | 7.1. Secuenciación do xenoma completo (Shotgun) 7.2. Secuenciación do xenoma bacteriano 16S rRNA 7.2.1. Anotación baseada en asignación de OTUs e ASVs 7.2.2. Análise de diferenzas en equilibrio microbiano 7.3. American Gut Project (AGP), Human Microbiome Project (HMP) |



| Planning | | | | |
|--------------------------------|--------------------------------|----------------------|-------------------------------|-------------|
| Methodologies / tests | Competencies | Ordinary class hours | Student?s personal work hours | Total hours |
| Laboratory practice | A2 B2 B9 B10 C1 C4 | 16 | 16 | 32 |
| Multiple-choice questions | B7 B8 | 3 | 21 | 24 |
| Supervised projects | A2 B2 B3 B4 B7 B9 B10 C1 C4 | 6 | 30 | 36 |
| Guest lecture / keynote speech | B7 B8 C1 C4 | 21 | 36 | 57 |
| Personalized attention | | 1 | 0 | 1 |

(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

| Methodologies | |
|--------------------------------|---|
| Methodologies | Description |
| Laboratory practice | Resolución de problemas e diferentes partes individuais do pipeline de análise de datos ómicos |
| Multiple-choice questions | Proba de avaliación escrita na que o/a estudante deberá demostrar os coñecementos adquiridos |
| Supervised projects | Elaboración supervisada de análise de datos ómicos reais mediante técnicas aprendidas ao longo das sesións maxistras e de prácticas de laboratorio da materia |
| Guest lecture / keynote speech | Impartición teórica da materia. Alternaranse a exposición de novos conceptos teóricos coa revisión de exemplos e a resolución de pequenos exercicios |

| Personalized attention | |
|--|--|
| Methodologies | Description |
| Laboratory practice Supervised projects | O profesor guiará o proceso de resolución dos problemas e liquidará as dúbidas que xurdan no desenvolvemento |

| Assessment | | | |
|---------------------------|--------------------------------|---|---------------|
| Methodologies | Competencies | Description | Qualification |
| Laboratory practice | A2 B2 B9 B10 C1 C4 | As/os estudantes deberán realizar pequenas partes dos pipelines de análises de datos ómicos para asentar o coñecemento teórico e prepararse para o traballo tutelado | 10 |
| Supervised projects | A2 B2 B3 B4 B7 B9 B10 C1 C4 | As/os estudantes deberán analizar datos ómicos reais aplicando as técnicas/aproximacións aprendidas na materia. Deberá acompañarse dun informe replicable de resultados, código e conclusións realizado en contorna R xustificando o pipeline de análise utilizada e as conclusións obtidas | 60 |
| Multiple-choice questions | B7 B8 | No período de avaliación realizarase unha proba de coñecementos teórico-prácticos para avaliar a adquisición das competencias por parte do alumno | 30 |

| Assessment comments |
|---------------------|
| |



A/o estudante deberá obter unha nota mínima de 3 sobre 10 puntos na proba obxectiva e unha nota mínima de 4 sobre 10 puntos nos traballos tutelados. As prácticas de laboratorio non teñen nota mínima. De non alcanzar esta nota mínima nalgún dos devanditos apartados, a nota da materia será a menor das dúas. As entregas correspondentes aos traballos tutelados nas datas indicadas e a asistencia ás clases prácticas son obrigatorias para aprobar a materia. Un/unha estudante considerárase presentado/a en unha convocatoria se fai entrega dos traballos tutelados ou se se presenta á proba obxectiva. O traballo entregado deberá ser orixinal do/da estudante. De acordo ao artigo 14, apartado 4, da normativa*, a entrega de traballos non orixinais ou con partes duplicadas (sexa por copias entre compañeiros ou por obtención doutras fontes...) levará unha nota global de SUSPENSO na convocatoria correspondente, tanto para o/a estudante que presente material copiado como a quen o facilitara, invalidando calquera outra cualificación obtida nas actividades avaliables.* Normativa de avaliación, revisión e reclamación das cualificacións dos estudos de grao e máster universitario, aprobada polo Consello de Goberno da Universidade da Coruña o 19 de decembro de 2013. Sobre a responsabilidade compartida dos traballos en grupo. Nas actividades que levan a cabo en grupos, tales como os traballos tutelados, todos os membros do grupo serán responsables solidarios do traballo realizado e entregado, así como das consecuencias que se deriven do incumprimento das normas de autoría do mesmo. Segunda oportunidade e convocatorias posteriores Na segunda oportunidade, mantense a nota obtida nos traballos tutelados. Só no caso de non presentar os traballos tutelados na primeira oportunidade ou que estas obtivesen a cualificación de SUSPENSO (cualificación menor que 5), poderanse presentar novamente os traballos tutelados na segunda oportunidade. En caso de suspender a materia, os traballos tutelados con nota igual ou superior a 5 gardaranse para cursos posteriores con cualificación de aprobado (5). En cada curso, o/a estudante terá a opción de entregar un novo traballo tutelado que substituirá a nota da anterior. Os traballos non se gardarán máis dun curso. Matrícula con dispensa académica Para estudantes con matrícula con dispensa académica elimínase a obrigatoriedade de asistencia ás clases de prácticas. Deberán entregar os traballos tutelados nas datas establecidas. É responsabilidade de devanditos/as estudantes poñer en coñecemento do profesor a súa circunstancia. Tutorías A maiores, as tutorías considéranse unha parte importante dentro do desenvolvemento da materia. Están orientadas de tal maneira que as/os estudantes teñan e/o poidan consultar distintas cuestións como: 1. Problemas no desenvolvemento das prácticas 2. Maneiras de enfocar/organizar os traballos tutelados 3. Resolución de dúbidas sobre as cuestións teóricas. Pedirase ás/os estudantes que soliciten cita aos/as profesores/as responsables para realizar reunións presenciais dentro dos horarios de tutorías do profesorado establecido en espazos.udc.es. O uso de Teams deberá limitarse a cuestións puntuais organizativas.

Sources of information

| | |
|----------------------|---|
| Basic | <ul style="list-style-type: none"> - Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. <i>Genome Biology</i> - Chen Y, Lun AAT, Smyth GK (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. <i>F1000Research</i> - Enis Afgan, Dannon Baker, Bérénice Batut, Marius van den Beek, Dave Bouvier, Martin ?ech, John Chilt (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. <i>Nucleic Acids Research</i> - TCGA Consortium (2022). The Cancer Genome Atlas. https://portal.gdc.cancer.gov/ - NCBI Gene Expression Omnibus (2022). NCBI Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/ - Michael Love, Wolfgang Huber y Simon Anders. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. <i>Genome Biology</i> - Malachi Griffith y col. (2015). Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. <i>Plos Computational Biology</i> |
| Complementary | <ul style="list-style-type: none"> - Liñares-Blanco J., Fernandez-Lozano C., Seoane JA y López-Campos G. (2022). Machine Learning Based Microbiome Signature to Predict Inflammatory Bowel Disease Subtypes. <i>Frontiers in Microbiology</i> - Fernández-Edreira D., Liñares-Blanco J. y Fernandez-Lozano C. (2021). Machine Learning analysis of the human infant gut microbiome identifies influential species in type 1 diabetes. <i>Expert Systems with Applications</i> - Liñares-Blanco, J., Gestal, M., Dorado, J., y Fernandez-Lozano, C. (2019). Differential gene expression analysis of RNA-seq data using machine learning for cancer research. <i>Machine Learning Paradigms. Learning and Analytics in Intelligent Systems. Vol 1. Springer, Cham.</i> |

Recommendations

Subjects that it is recommended to have taken before



Statistical Analysis of Complex Data/614G02031
Machine Learning I/614G02019
Multivariable Calculus /614G02006

Subjects that are recommended to be taken simultaneously

Subjects that continue the syllabus

Statistical Analysis of Complex Data/614G02031
Large Scale Machine Learning/614G02032

Other comments

- Segundo se recolle nas distintas normativas de aplicación para a docencia universitaria deberase incorporar a perspectiva de xénero nesta materia (usarase linguaxe non sexista, utilizarase bibliografía de autores/as de ambos sexos, propiciarse a intervención en clase de alumnos e alumnas...)- Traballarase para identificar e modificar prexuizos e actitudes sexistas e influirase na contorna para modificalos e fomentar valores de respecto e igualdade.- Deberanse detectar situacións de discriminación por razón de xénero e proponerse accións e medidas para corrixilas.

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.